ASSESSING THE INFLUENCE OF PACKET LOSS AND FRAME FREEZES ON THE PERCEPTUAL QUALITY OF FULL LENGTH MOVIES

Nicolas Staelens\textsuperscript{a} Brecht Vermeulen\textsuperscript{a} Stefaan Moens\textsuperscript{b} Jean-François Macq\textsuperscript{c}
Peter Lambert\textsuperscript{b} Rik Van de Walle\textsuperscript{b} Piet Demeester\textsuperscript{a}
\textsuperscript{a}Ghent University - IBBT, Department of Information Technology (INTEC), Ghent, Belgium
\textsuperscript{b}Ghent University - IBBT, Department of Electronics and Information Systems (ELIS), Ghent, Belgium
\textsuperscript{c}Alcatel-Lucent, Bell Labs - Fixed Access - Video Technologies, Antwerp, Belgium

ABSTRACT
Since the introduction of digital television, new services such as Video on Demand and IPTV have been widely deployed. These new enhanced video services are usually delivered via streaming technologies over IP-based networks to set-top boxes (STB). Unfortunately, such networks are commonly referred to as packet-based best-effort networks that cannot guarantee the correct delivery of the transmitted data. As such, network impairments like packet loss can cause severe visual degradations during video playback. Current existing methodologies for the subjective quality assessment of video sequences pose tight restrictions on viewing conditions and experiment duration. As a consequence, these methodologies cannot be used for the quality assessment of full length movies in real-life conditions. In this paper, a new methodology is proposed and used to evaluate the influence of packet loss and frame freezes on the perceived visual quality of full length movies.

1. INTRODUCTION
The introduction of digital television changed the face of how users experience and consume media. Several new enhanced video services such as Video on Demand (VoD) and Internet Protocol Television (IPTV) have been widely deployed, enabling users to watch according to their own schedules and preferences. VoD enables viewers to buy (via a pay-per-view mechanism) or rent a movie, a television program or any other available content any time they want it and as often as they want. Furthermore, it is generally known that watching television is a typical lean-backward experience. These new enhanced video services are usually delivered via streaming technologies over IP-based networks to STBs. Unfortunately such networks, like the Internet, are commonly referred to as packet-based best-effort networks that cannot guarantee the correct delivery of the transmitted data. As such, network impairments like packet loss, delay and jitter can cause severe visual degradations during video playback \cite{1, 2, 3, 4, 5}.

In order to assess the visual quality of video sequences, several subjective video quality assessment methodologies have already been standardized \cite{6, 7, 8} which describe in detail how subjective video assessment tests should be organized and conducted. These methodologies specify amongst others how the video sequences must be presented to the subjects and pose limits on the maximum durations of both the video sequences and the entire test. As such, a typical subjective video quality assessment test should not take more than 30 minutes in order to avoid viewer fatigue and the length of each video sequence which must be evaluated by the subjects should be limited to 15 seconds. Furthermore, these international standardized methodologies also describe the viewing conditions under which the test must be carried out. These viewing conditions, in turn, specify the viewing distance between the screen and the human observer, the peak luminance level of the screen and the screen contrast ratio. As a consequence, subjective tests are usually expensive and very time consuming.

Watching television is a typical lean-backward experience. Unfortunately, current existing subjective test methodologies do not incorporate this lean-backward television viewing experience since human subjects are usually instructed to actively evaluate visual quality. During subjective tests, human subjects are therefore searching for visual impairments and visual degradations. Furthermore, it is clear that the limitations concerning the entire duration of the subjective tests, the duration of the individual video sequences and the stringent demands on the viewing conditions make the existing methodologies inappropriate for assessing the visual quality of full-length movies in real-life environments.

In this paper, a new methodology is proposed to assess the perceived visual quality of full length movies in real-life conditions. Afterwards, this new subjective methodology is used to evaluate the influence of packet loss and frame freezes on the perceptual quality of full length movies.
The rest of this paper is organized as follows: section 2 describes related work concerning subjective video quality assessment methodologies and already performed studies. In section 3, a new subjective methodology is proposed to assess the visual quality of full-length movies in real-life conditions. Section 4 describes the conducted subjective test for assessing the influence of packet loss and frame freezes in full-length movies. The results of this study are presented in section 5.

2. RELATED WORK

In [9], the influence of primacy and recency effects on overall quality rating of video sequences is studied. Primacy and recency effects occur when the earliest, respectively the latest information plays a more important role on the judgement. In the case of video quality evaluation, these order effects correspond with visual impairments occurring either in the beginning or at the end of the video sequence. For the experiment, 30s sequences were used and presented to the subjects using the Double Stimulus Continuous Quality Scale (DSCQS) [6]. Results indicate a significant difference between the quality ratings of sequences which contain an impairment at the beginning compared to sequences which contain an impairment at the end.

The effect of sporadically dropped frames on overall perceived quality is studied in [10]. Subjective tests were conducted with video sequences of 10 seconds long to identify the perceptual detection threshold of a single temporal discontinuity (frame freeze) and to quantify the effect of several impairments distributed over time. Results revealed that a single frame freeze with a duration of 80ms was noticed by 80% of the subjects. When this duration was increased to 200ms, 100% of the subjects perceived the impairment.

International subjective video quality evaluation methodologies as specified in [6] and [7] include guidelines on how to perform a variety of subjective tests. These tests are either Single Stimulus tests, where only one video sequence at a time is shown, or Double Stimulus tests, where video sequences are shown pairwise. Next to defining the order in which the sequences are presented, these international standards also specify the grading scales that subjects must use in order to rate the quality of the video sequences. Two commonly used grading scales are the Absolute Category Rating (ACR) scale and the Continuous Quality scale. In the former, an evaluation is performed after a video sequence has been watched; in the latter, users can rate the sequences while they are still playing. Another important aspect of the subjective quality methods is that they specify that maximum durations of the video sequences that must be evaluated. As such, the length of the video sequences should not exceed 15 seconds.

The Subjective Assessment Methodology for Video Quality (SAMVIQ) [8] is based on the Double Stimulus Continuous Quality Scale (DSCQS) from [7] but is more efficient in the assessment of a larger range of image qualities. The SAMVIQ methodology differs from the DSCQS methodology in the fact that it is able to discriminate low qualities as well as high qualities. Furthermore, the test subjects can start or stop the evaluation and change their quality score for each of the sequences as often as they want. It is believed that this subjective quality assessment method is appropriate for the evaluation of multimedia content, specifically for the evaluation of multimedia codecs [11]. Since the SAMVIQ methodology is based on the DSCQS methodology, only short video sequences (typically between 10 seconds and 15 seconds) should be evaluated. The SAMVIQ methodology uses a multi-stimulus form that enables users to choose on their own the order of the test and enables them to re-evaluate previous sequences. By this, users are able to grade each sequence accordingly and thus discriminate between the different levels of quality [12].

3. FULL LENGTH MOVIE QUALITY ASSESSMENT METHODOLOGY

In order to evaluate the perceived visual quality of full-length movies in real-life environments, a new subjective methodology is needed which:

- enables subjects to watch the video sequence in their most natural environment (e.g. in their living room)
- ensures no additional visual impairments¹ are injected during video playback
- does not encourage human subjects to focus on actively evaluating visual quality

To satisfy the above mentioned requirements, we propose the use of full-length DVD movies. This approach enables us to provide a DVD, that contains an impaired full-length movie, to the subjects which can be taken home. We hereby encourage subjects to watch the DVD in the same environment they also watch their rented or bought DVDs. Furthermore, by creating compliant DVDs, no additional visual impairments will be injected during movie playback.

Fig. 1 presents the toolchain we used to create impaired full-length DVD movies². In the first steps, the video and audio tracks are extracted from the main movie of the DVD. Once these tracks have been demuxed, impairments can be injected in the video and/or audio track. After the creation of the impaired tracks, audio and video must be remuxed and written onto a new DVD.

¹By this we mean visual impairments beside the ones injected intentionally in the video sequence.
²We strongly emphasize that this toolchain was used to created DVDs for research purposes only.
4. SUBJECTIVE TEST FOR THE QUALITY ASSESSMENT OF FULL LENGTH MOVIES

For the subjective test, seven full length movies of different genres (action, adventure, comedy, drama) were used. The movies were impaired with packet loss and frame freezes using the toolchain from the previous section. It must be pointed out that only the video track of the movie was impaired while the audio track remained intact. As such, audio quality remained constant during movie playback. Each movie was impaired with up to three visual impairments, spread over time. By this, full length DVD movies were created with:

- a visual impairment in the first and/or last half hour of video playback
- a visual impairment in the first and/or last half hour and a frame freeze in the middle of the movie.

The visual impairment in the first and/or last half hour was either a single frame freeze or random blockiness caused by packet loss. Packet loss rates were selected so that about 50% of each image in the impaired fragment contained blockiness. The duration of the impairments varied between 320ms and 400ms. The impaired movies were then burned onto a DVD disc that could be taken home by the subjects.

Subjects were asked to complete a questionnaire immediately after watching the DVD but they were not informed in advance about any possible visual degradations during playback. One of the questions they needed to answer was how many and which type of errors they had observed. They were also asked to describe the scene within the movie that contained the specified errors. The subjects also were required to provide a Mean Opinion Score (MOS), using a 5-grade quality scale with adjectives, for the overall quality of the movie. Finally, subjects were also asked to describe the typical viewing conditions under which they watched the DVD. This included questions concerning the type of television on which they watched the movie and the viewing distance between the TV-set and the position of the subject. They were also asked to describe the lighting conditions.

5. RESULTS

A total number of 56 non-expert viewers (32 males and 24 females), aged between 14 and 49 years, participated in the subjective test. All subjects together watched 80 DVDs.

5.1. Typical home viewing conditions

Nineteen subjects were asked to describe their typical home viewing conditions. In 58% of the cases, the DVD was watched on a CRT television set; 42% of the subjects had a LCD television. Table 1 summarizes the main findings concerning the television set on which the DVD was watched.

<table>
<thead>
<tr>
<th>Type</th>
<th>Aspect ratio</th>
<th>Screen Size (in inches)</th>
<th>Viewing distance (in screen heights)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRT</td>
<td>4:3</td>
<td>17 - 27</td>
<td>5H - 9H</td>
</tr>
<tr>
<td>LCD</td>
<td>16:9</td>
<td>23 - 40</td>
<td>5H - 10H</td>
</tr>
</tbody>
</table>

Table 1. Subject’s home television set characteristics.

All CRT screens had a 4:3 aspect ratio with a size varying from 17 to 27 inches; LCD television screens were all widescreen (16:9), with sizes between 23 and 40 inches. The viewing distance, expressed in terms of the screen height, varied from 5H to 9H and from 5H to 10H for respectively CRT and LCD screens. All DVDs were watched in the living room during the evening.

5.2. Influence of frame freezes and blockiness on perceived visual quality

In order to validate the recency effect from [9], MOS scores from subjects who perceived only one visual impairment in the first half hour of the movie were compared with the MOS scores from viewers who perceived the same impairment in the last half hour of the movie. A t-test was used to check the significance level. Table 2 lists the different p-values for the different types of perceived impairments.

<table>
<thead>
<tr>
<th>Impairment</th>
<th>t</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame freeze</td>
<td>0.7108</td>
<td>22.991</td>
<td>0.4843</td>
</tr>
<tr>
<td>Blockiness</td>
<td>1.7883</td>
<td>20.17</td>
<td>0.08876</td>
</tr>
</tbody>
</table>

Table 2. Calculated p-values for testing significance levels between MOS scores of perceived visual impairments in the first and last half hour of movie playback.

Looking at the p-values for frame freezes and blockiness it can be concluded that the location of the visual impairment, being either in the first or last half hour of the movie, does not significantly influence overall quality rating. This means that there is no recency effect when one visual impairment is observed during video playback.

In general, frame freezes are less noticeable than blockiness, as illustrated in Fig. 2(a) and 2(b).
Only 34% of the viewers who evaluated a DVD containing a frame freeze in the first half hour actually perceived the impairment. In the case of blockiness all viewers noticed the impairment in the first half hour. When the impairment occurred during the last half hour of movie playback, 50% of the subjects perceived the frame freeze compared to 81% in case of blockiness. Since frame freezes are much less noticeable than blockiness, they can be used as a simple yet effective error concealment technique.

Subjects were asked to indicate which visual impairment appeared the most and the least annoying when they perceived both frame freezes and blockiness. Twelve viewers responded to this question. From the results in Fig. 3 it is clear that subjects rated frame freezes being more annoying than blockiness.

Fig. 4 shows the average MOS scores based on the number of perceived visual impairments. On average, the overall quality rating was 4.52 in case no impairments were observed during the entire movie. The MOS score from subjects who perceived one and two impairments were respectively 4.28 and 4.29. In these cases, the overall quality score was independent of the location and the type of the visual degradation.

To validate whether MOS scores were significantly influenced by the number of perceived visual errors, p-values were calculated using a t-test. Significance levels were tested between:

- MOS scores from subjects who perceived no impairments and MOS scores from subjects who perceived one impairment
- MOS scores from subjects who perceived no impairments and MOS scores from subjects who perceived two impairments
- MOS scores from subjects who perceived one impairment and MOS scores from subjects who perceived two impairments

From the values in Table 3 it can be concluded that there is no significant difference between the MOS scores of subjects who perceived zero, one or two visual impairments. In other words, subjects perceiving up to two visual impairments did not rate quality worse compared to subjects who did not perceive any impairments at all.
6. CONCLUSIONS

In this paper, a new subjective methodology was proposed to assess the perceptual influence of frame freezes and blockiness in full length movies. Full length DVD movies were used which contained some visual degradations distributed over time. Using this approach, the DVDs could be taken home by the subjects and watched in a real-life environment.

Subjective tests were conducted to evaluate the influence of frame freezes and packet loss on the quality assessment of full length movies. Results were compared with earlier studies conducted with international standardized subjective quality assessment methods. Our tests revealed that subjects are more tolerant towards visual impairments when they are not mainly focused on quality evaluation.

7. ACKNOWLEDGEMENTS

The research activities that have been described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT) and the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT). Nicolas Staelens would like to thank the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT) for financial support through his Ph.D. grant. Part of this work has been funded by the IBBT GBO project Video Q-SAC.

8. REFERENCES


